

Risk charts to guide targeted HIV-1 viral load monitoring of ART: development and validation in patients from resource-limited settings

Manuel Koller, PhD^a, Geoffrey Fatti, MD^b, Benjamin H Chi, MD^c, Olivia Keiser, PhD^a, Christopher J Hoffmann, MD^d, Robin Wood, MD^e, Hans Prozesky, MD^f, Kathryn Stinson, MD^g, Janet Giddy, MD^h, Portia Mutevedzi, PhDⁱ, Matthew Fox, PhD^{j,k}, Matthew Law, PhD^l, Andrew Boulle, MD, PhD^m, and Matthias Egger, MD^{a,m}

on behalf of the International epidemiological Databases to Evaluate AIDS in Southern Africa (IeDEA-SA) and the TREAT Asia HIV Observational Database (TAHOD)

- a) Institute of Social and Preventive Medicine, University of Bern, Switzerland
- b) Kheth'Impilo, Cape Town, South Africa
- c) Centre for Infectious Disease Research in Zambia, Lusaka, Zambia
- d) Aurum Institute for Health Research, Johannesburg, South Africa
- e) Gugulethu ART Programme and Desmond Tutu HIV Centre, University of Cape Town, South Africa
- f) Division of Infectious Diseases, Department of Medicine, University of Stellenbosch and Tygerberg Academic Hospital, Cape Town, South Africa
- g) Médecins Sans Frontières, Khayelitsha, Cape Town, South Africa
- h) Sinikithemba Clinic, McCord Hospital, Durban, South Africa
- i) Africa Centre for Health and Population Studies, University of KwaZulu-Natal, Somkhele, South Africa
- j) Health Economics and Epidemiology Research Office, University of the Witwatersrand, Johannesburg, South Africa
- k) Center for Global Health & Development and Department of Epidemiology, Boston University, Boston, USA
- l) Biostatistics and Databases Program, The Kirby Institute, Faculty of Medicine, The University of New South Wales, Sydney, Australia
- m) Centre for Infectious Disease Epidemiology and Research, School of Public Health and Family Medicine, University of Cape Town, South Africa

Correspondence to:
Professor Matthias Egger
Institute of Social and Preventive Medicine (ISPM)
University of Bern
Finkenhubelweg 11, CH-3012 Bern, Switzerland
egger@ispm.unibe.ch

Abstract 242 words, main text 3306 words, 2 tables, 3 figures, 39 references, supplementary web appendix (with 6 tables, 3 figures, and a technical appendix)

Competing Interests: The authors have declared that no competing interests exist.

Sources of Funding: IeDEA Southern Africa and TAHOD are supported by the National Institutes of Health (NIH): National Institute of Allergy and Infectious Diseases (NIAID), National Institute of Child Health and Human Development (NICHD), the Office of the Director (OD), and the National Cancer Institute (NCI), as part of the International Epidemiologic Databases to Evaluate AIDS (IeDEA, grants U01AI069924 and U01AI069907). TAHOD is part of the Asia Pacific HIV Observational Database and is an initiative of TREAT Asia, a program of amfAR and The Foundation for AIDS Research. TAHOD is also supported by the Dutch Ministry of Foreign Affairs through a partnership with Stichting Aids Fonds. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. M Fox was supported by Cooperative Agreement AID 674-A-12-00029 from the United States Agency for International Development (USAID). OK was supported by a PROSPER Ambizione fellowship of the Swiss National Science Foundation.

Keywords: Antiretroviral therapy; monitoring; CD4 cell counts; risk charts; tripartite decision rules; Sub-Saharan Africa

Abstract

Background: HIV-1 RNA viral load (VL) testing is recommended to monitor antiretroviral therapy (ART) but not available in many resource-limited settings. We developed and validated CD4-based risk charts to guide targeted VL testing.

Methods: We modeled the probability of virologic failure up to 5 years of ART based on current and baseline CD4 counts, developed decision rules for targeted VL testing of 10%, 20% or 40% of patients in seven cohorts of patients starting ART in South Africa, and plotted cut-offs for VL testing on colour-coded risk charts. We assessed the accuracy of risk chart-guided VL testing to detect virologic failure in validation cohorts from South Africa, Zambia and the Asia-Pacific.

Findings: 31,450 adult patients were included in the derivation and 25,294 patients in the validation cohorts. Positive predictive values increased with the percentage of patients tested: from 79% (10% tested) to 98% (40% tested) in the South African, from 64% to 93% in the Zambian and from 73% to 96% in the Asia-Pacific cohorts. Corresponding increases in sensitivity were from 35% to 68% in South Africa, from 55% to 82% in Zambia and from 37% to 71% in Asia-Pacific. The area under the receiver-operating curve increased from 0.75 to 0.91 in South Africa, from 0.76 to 0.91 in Zambia and from 0.77 to 0.92 in Asia Pacific.

Interpretation: CD4-based risk charts with optimal cut-offs for targeted VL testing may be useful to monitor ART in settings where VL capacity is limited.

Introduction

Since 2002 the number of HIV-positive people receiving antiretroviral therapy (ART) in low- and middle-income countries has increased dramatically, from 300,000 in 2002 to 10 million by the end of 2012, representing two thirds of the United Nations target of 15 million people on ART by 2015.¹ The massive scale-up of ART also increased the number of patients experiencing treatment failure, the need for more expensive second-line regimens, and levels of viral resistance.^{2,3}

Clinical and laboratory monitoring of patients on ART aims to maximize the durability of first-line regimens. In high-income countries plasma HIV 1-RNA viral load (VL) and CD4 positive T cell count (CD4 count) are regularly measured and tests are done when drug resistance is suspected.⁴ In resource-limited settings monitoring of ART is, however, still generally based on CD4 counts and signs and symptoms. The accuracy of the criteria proposed by the World Health Organization (WHO)⁵ to detect virologic failure based on CD4 count and clinical criteria is poor: the positive predictive value (PPV) and sensitivity are below 50%.^{6,7} Patients with suppressed viral replication may thus unnecessarily be switched to second-line ART, and patients who fail therapy will switch late, or not switch at all.⁸ The 2013 WHO consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection and the March 2014 supplement recommend routine VL monitoring, but recognize that scaling up VL testing in resource limited settings will be challenging.^{9,10}

In settings where VL is not monitored routinely the first priority should be to confirm virologic failure in patients in whom treatment failure is suspected, based on CD4 count and clinical monitoring.¹⁰ Targeted VL testing of selected patients based on CD4 count and other criteria is promising in this situation: only relatively few patients have to be tested, thus reducing costs compared to routine VL monitoring.¹¹ We developed and validated risk charts based on current and past CD4 counts and decision rules to guide targeted VL testing.

Methods

Data sources

The International epidemiologic Databases to Evaluate AIDS in Southern Africa (IeDEA-SA) is a regional collaboration of HIV treatment and care programs, which is part of a consortium of seven networks in sub-Saharan Africa, Asia and Pacific, North America, and Caribbean, Central and South America.^{12–14} Data are collected at ART initiation and each follow-up visit, using standardized

instruments, and transferred in regular intervals to data centers in Switzerland and South Africa. Ethical approval was obtained from the ethics committee of the Canton of Bern, Switzerland and the University of Cape Town, South Africa. All participating cohorts obtained local ethical committee approval to contribute data to this analysis.

We developed the risk charts using seven South African cohorts: the Gugulethu and Khayelitsha township ART programs and Tygerberg hospital in Cape Town,^{15–17} the McCord Hospital in Durban,¹⁸ the Helen Joseph Hospital Themba Lethu Clinic and Aurum Institute for Health Research program in Johannesburg,^{19,20} and the Hlabisa HIV Treatment and Care program in rural Somkhele, KwaZulu-Natal.²¹ We describe the seven cohorts, which mainly include urban and township populations, as the derivation dataset. We validated the risk charts in the South African Kheth'Impilo cohort, which includes health facilities from urban and rural areas in the Eastern Cape, KwaZulu-Natal and Mpumalanga,²² the Centre for Infectious Diseases and Research in Zambia (CIDRZ), which covers urban and periurban populations in Lusaka,²³ and the TREAT Asia HIV Observational Database (TAHOD) in the Asia-Pacific.²⁴

In South Africa all cohorts monitored VL and CD4 cell counts six-monthly. Similarly, from TAHOD we included 17 sites in 12 countries that routinely monitored VL and CD4 counts three to six monthly. In CIDRZ monitoring of CD4 cell counts occurs every three to six months and VL is measured in patients suspected of failing therapy.

Inclusion criteria, definitions and imputations of missing values

We included treatment-naïve patients aged 16 years or older who started first-line ART in 2000 or later with a CD4 cell count of 350 cells/ μ L or lower. Patients needed to have at least one VL measurement and one CD4 count 6 months or later after starting ART. The CD4 cell count at the start of ART was defined as the measurement closest to the date of starting ART, within a window of 90 days prior to 30 days after the start of ART. We defined virologic failure as a single VL above 1000 copies/ml. We included measurements taken up to 5 years after starting ART. In both the derivation and validation cohorts we imputed values missing between two measurements by interpolating values on the \log_{10} scale for VL and the square root scale for CD4 count. Measurements taken after switching to second-line ART were excluded.

Development of risk charts and rules for targeted VL testing

We used generalized additive models²⁵ with a logit link and thin-plate regression splines²⁶ with a monotonicity constraint to model the probability of virologic failure and develop the risk charts. In model 1 we included the current CD4 count, the CD4 count at start of ART, time on treatment and gender. In model 2 the CD4 count at ART initiation was replaced by a count measured 6 months

earlier, within a window of 2 and 9 months earlier. Most patients contributed multiple measurements during follow-up; these were treated as independent. Models included smoothers for the current CD4 count, time on treatment and age. We developed optimal tripartite decision rules to support decisions on VL testing in settings where access to VL monitoring is limited, using a method developed by Liu et al.²⁷ A tripartite decision rule is defined by two cut-off values that classify treatment outcomes into three categories, based on the predicted probability of virologic failure: successful ART, virologic failure, and uncertain outcome. VL is then measured in patients with uncertain outcome. We developed decision rules assuming that resources allow for VL testing of 10%, 20% or 40% of patients. The cutoffs were then chosen such that the 10%, 20% or 40% of patients with the most uncertain outcome are tested. We also determined the optimal cut-off in the absence of VL testing. In all rules we gave more weight to avoiding false negatives (60%) than to avoiding false positives (40%). In other words we assumed that it is more important to avoid missing patients who truly failed than to avoid falsely classifying patients as failing treatment.

Validation of risk charts and model fit

We calculated positive predictive values (PPV), negative predictive values (NPV), sensitivity, specificity and the area under the receiver-operating curve (AUC) for the derivation and validation cohorts. We checked the goodness of fit of the two models by graphically comparing observed and predicted risks. We overlaid the plots with a grid of 15x15 cells and compared the proportion of failures encountered within each of the 225 cells with the predicted number of failures. Finally, we compared the performance of the risk charts with the 2006 and 2013 WHO immunologic criteria for treatment failure.^{5,9} The 2006 criteria include a fall of the CD4 count to baseline (or below), a 50% fall from the on-treatment peak value and persistent CD4 counts below 100 cells/ μ l. The 2013 criteria are a simplified version of the 2006 criteria that do not include the fall from the on-treatment peak value.

Sensitivity analyses

We performed three sensitivity analyses. The first was a complete case analysis for which we did not impute any missing values. In the second sensitivity analysis we used an alternative imputation method where we added random error to interpolated values. Finally, in the third sensitivity analysis we examined the impact of assuming that multiple measurements in the same patient are independent. We weighted each measurement such that the weights of all measurements of one patient added up to 1. Every patient thus contributed the same weight to the analysis. See technical appendix, Supplemental Digital Content, <http://links.lww.com/QAI/A712> for further details. All analyses were done in R 3.0.1 (R Core Team, Vienna, Austria).

Results

Selection of eligible patients

After excluding patients with missing or ineligible CD4 counts at the start of ART, 31,450 patients from the South African derivation cohort, 16,131 patients from the South African, 7,796 patients from the Zambian and 1,356 from the Asia Pacific validation cohorts were included in the development and validation of the risk charts based on model 1. Numbers were different for the second risk chart (model 2), which was based on current and CD4 counts measured 6 months previously: 36,511 patients from the derivation cohort in South Africa and 12,909 patients from the South African, 2,854 patients from the Zambian and 1,367 patients from the Asia-Pacific validation cohorts. The selection of patients with reasons for exclusion is shown in Figure S1, Supplemental Digital Content, <http://links.lww.com/QAI/A712>.

Patient characteristics

The 31,450 patients starting ART in one of the seven South African programs had a median age of 36 years, were predominantly female (18,597; 59%) and started ART with a median CD4 cell count of 111 cells/ μ L (Table 1). The characteristics of the patients included in the South African and Zambian validation cohorts were similar to the derivation cohort, whereas in the cohorts from Asia-Pacific, most patients were men (918; 68%) and the median CD4 cell count at start of ART was lower, 95 cells/ μ L. For model 2, patient characteristics were similar (see Table S1, Supplemental Digital Content, <http://links.lww.com/QAI/A712>).

The development of model 1 was based on 125,590 triplets of laboratory values: the CD4 count measured at the start of ART and one CD4 count and VL measured subsequently at the same point in time during a total 68,611 person-years of follow-up. The validation datasets were based on 46,997 triplets measured during 32,005 person-years of follow-up (South Africa), 16,652 triplets measured during 19,951 person-years (Zambia) and 8,498 triplets taken during 4,375 person-years (Asia-Pacific). In the derivation cohorts 11,972 (10%) of VL values and 12,045 (10%) of CD4 counts had been imputed by interpolation. Compared to the derivation cohorts the proportion of imputed values was greater in the validation cohorts from South Africa and Zambia, and greater for VL in Asia-Pacific (Table 1). The numbers were similar for model 2 (see Table S1, Supplemental Digital Content, <http://links.lww.com/QAI/A712>).

Risk charts

The risk chart for virologic failure based on model 1 is shown in Figure 1, stratified by CD4 count at the start of ART and gender. Low probabilities of virologic failures are shown in blue, intermediate probabilities in yellow and orange, and high probabilities in red. At a given combination of current and CD4 count at the start of ART, the probability of virologic failure increases with time on ART, and is somewhat lower in women than in men. The optimal probability area where patients should be tested if resources allow the testing of 10%, 20% or 40% of patients are also shown. The range of patients to be tested widens with duration on ART, reflecting increasing uncertainty. Figure 2 shows the risk chart for model 2, stratified by CD4 count measured 6 months previously and gender. Again, at a given combination of current and previous CD4 count, the probability of virologic failure increases with time on ART, and is lower in women than in men. Alternative presentations of the two risk charts are given in Figure S2 and Figure S3, Supplemental Digital Content, <http://links.lww.com/QAI/A712>.

Accuracy of charts and targeted VL testing

The range of probabilities resulting in the testing of 10%, 20% and 40% of patients in the derivation cohort slightly differed between models. E.g., when assuming that 20% of patients can be tested this range was 0.22 to 0.64 for model 1 and 0.20 to 0.67 for model 2. With model 1 the PPV increased from 61% to 87%, 94% and 98% in the South African derivation cohort when moving from no VL testing to the testing of 10%, 20% and 40% of patients (Table 2). The PPVs for the South African validation cohort increased from 48% (no testing) to 79% (10% tested), 91% (20% tested) and 98% (40% tested). The corresponding PPVs for Zambia were 35%, 64%, 80% and 93%, and for Asia-Pacific 37%, 73%, 88% and 96%. NPVs were close to 90% in all cohorts even without targeted VL testing, and generally above 90% with targeted testing, except for the South African validation cohort with no testing (81%) and testing of 10% and 20% of patients (84%; 87%). Sensitivities increased from 33% (no VL testing) to 74% (40% tested) in the derivation cohort and, in the validation cohorts, from 24% (no VL testing) to 68% (40% tested) in South Africa, from 43% to 82% in Zambia and from 25% to 71% in the Asia-Pacific cohorts (Table 2). The AUCs ranged from 0.63 using model 2 in the Zambian validation cohort without targeted VL testing to 0.95 in the South African derivation cohorts when using model 1 and assuming that 40% of patients had VL tests (Figure 3).

Table S2, Supplemental Digital Content, <http://links.lww.com/QAI/A712> gives PPVs, NPVs, sensitivity and specificity for different threshold probabilities for virologic failure, assuming that targeted VL testing is not available. As expected, the PPV increased with higher thresholds whereas sensitivity declined. The comparison with the WHO criteria for immunological failure showed that in the absence of targeted VL monitoring the performance of the risk charts and the different WHO criteria

was similar (see Table S3, Supplemental Digital Content, <http://links.lww.com/QAI/A712>). The goodness of fit of the two models, as assessed by comparing observed and predicted risks, was generally high (see technical appendix, Supplemental Digital Content, <http://links.lww.com/QAI/A712> for details).

Sensitivity analyses

The results of the complete case analysis were similar to the main analysis, with only small differences in the accuracy of predictions, typically in the range of plus or minus 0% to 5% in PPV, NPV, sensitivity or specificity (see Table S4, Supplemental Digital Content, <http://links.lww.com/QAI/A712>). The same was the case when adding random normal errors to the imputed values used in the main analysis (see Table S5, Supplemental Digital Content, <http://links.lww.com/QAI/A712>). Finally, the results of the analysis to which each patient contributed the same weight were also similar to the main analysis (see Table S6, Supplemental Digital Content, <http://links.lww.com/QAI/A712>).

Discussion

The measurement of the CD4 count remains necessary to assess ART eligibility in many settings, and the CD4 count is also important to gauge the risk of clinical progression and guide clinical decisions about prophylactic treatments and screening for opportunistic infections.^{9,10} We used data from the large IeDEA collaboration to develop and validate charts of the risk of virologic failure in adult HIV-positive patients starting ART, based on two CD4 counts measured at different points in time. Over 30,000 adult patients starting ART were involved in the development of the charts, and up to 25,000 patients were included in their validation. The risk charts define optimal ranges of risk at which patients should be tested for VL, assuming that resources permitted the targeted testing of 10%, 20%, or 40% of patients. The PPVs increased substantially with targeted VL testing, even when only 10% of patients were tested, and was around 90% with the testing of 20% of patients. Sensitivity also increased: the decision rule based on the testing of 20% of patients identified 50% to 70% of patients with virologic failure.

The development of the charts based on seven South African urban and townships cohorts, the definition of tripartite decision rules using state-of-the-art methods,²⁷ and the thorough validation are important strengths of this study. The risk charts were validated in a large ART program in the Eastern Cape, KwaZulu-Natal and Mpumalanga provinces of South Africa, which included many rural treatment sites, a treatment program in the greater Lusaka metropolitan area in Zambia, and programs in 12 different countries in the Asia-Pacific region. Our study therefore meets several

dimensions of generalizability and applicability, including geographic, spectrum and methodological transportability.²⁸ Indeed, accuracy was maintained when the charts were tested in patients from different locations, with more or less advanced immunodeficiency, and across sites that differed with respect to data collection, follow-up intervals and monitoring strategies. Such multiple validations are possible only in large international cohort collaborations such as IeDEA.¹²

VL testing to monitor ART is strongly recommended by WHO.^{9,10} However, a 2012 survey found that few programs in Sub-Saharan Africa had access to routine VL testing.²⁹ As discussed in detail in recent WHO and UNAIDS reports,^{10,30} scaling up VL testing in resource limited settings is challenging. For example, plasma obtained from EDTA coagulated whole blood is the preferred sample for the common VL platforms, but obtaining plasma may not be feasible in remote clinics, due to the lack of electricity to operate centrifuges and maintain the cold-chain.¹⁰ Point-of-care laboratory tests are being developed both for CD4 cell count and VL.³¹ Some point-of-care VL tests are designed to be used in clinics in remote settings, by auxiliary staff and in the absence of a reliable electricity supply. The risk charts may facilitate the cost-effective use of point-of-care and standard VL tests,³² and generally support the transition to routine VL monitoring.¹⁰

As in previous studies^{33–35} we defined virologic failure as a single VL above 1000 copies/ml. Virologic failure should not be confused with treatment failure, which is defined by WHO as two consecutive VL measurements exceeding 1000 copies/ml, within a three-month interval, with adherence support between measurements, after at least six months of using ARV drugs.⁹ Also, we stress that the risk charts inform decisions on VL testing and adherence support but they do not on their own provide conclusive evidence for switching patients to second-line ART. Furthermore, although our study assessed the accuracy of the risk charts in different patient populations, it did not examine the effects of using these charts to monitor patients starting ART in resource-limited settings. Ideally, different monitoring strategies should be compared in pragmatic randomized trials with patient-relevant outcomes such as disease progression and mortality. Previous trials compared clinical monitoring with routine CD4 count monitoring, or CD4 count with CD4 count and VL monitoring.^{36,37} To our knowledge no trials of risk-based targeted VL monitoring have been done.

The models underlying the charts might be improved by including other variables predictive of virologic failure. The lack of data on adherence is an important limitation of our study: in the Aid for AIDS program in Southern Africa adherence assessments based on pharmacy refill data were as accurate as CD4 counts for detecting virologic failure.³⁸ A clinical prediction rule developed at the Sihanouk Hospital Center of Hope in Cambodia (based on adherence, and changes in CD4 cell count and hemoglobin values) had a sensitivity close to 50% and specificity of over 90%.³³ The performance of other scoring systems was similar, with improved sensitivity compared to the WHO criteria.^{34,35}

Few of these scores had undergone external validation, but it is noteworthy that the sensitivity of the Cambodian score dropped to 23% when used in Uganda.³⁴

Our study has other limitations. We only considered patients starting ART at CD4 cell counts of 350 cells/ μ L or below. Some countries are moving towards initiating patients at a CD4 count below 500 cells/ μ L and initiate ART in all pregnant women regardless of CD4. However, most patients still initiate ART at much lower CD4 counts. For example, in 2013, the median CD4 cell count was 231 cells/ μ L in the Republic of South Africa, 212 cells/ μ L in Malawi, 205 cells/ μ L in Botswana and 180 cells/ μ L in Tanzania.³⁹ The charts will therefore be relevant to many adult patients, and a similar study in children is now under way. Also, the charts will be updated and extended to beyond 5 years as more data accumulate in the leDEA cohorts.

In conclusion, the risk charts developed and validated in this study should be useful for a range of ART programs and settings, including programs that have relied on CD4 count monitoring and are now transitioning to targeted or routine VL testing. In settings that continue to have no access to VL testing the charts may provide a more user-friendly alternative to the WHO immunologic criteria for treatment failure.^{5,9} Field studies are now required to clarify the utility of these charts.

Acknowledgements: leDEA Southern Africa and TAHOD are supported by the National Institutes of Health (NIH): National Institute of Allergy and Infectious Diseases (NIAID), National Institute of Child Health and Human Development (NICHD), the Office of the Director (OD), and the National Cancer Institute (NCI), as part of the International Epidemiologic Databases to Evaluate AIDS (leDEA, grants U01AI069924 and U01AI069907). TAHOD is part of the Asia Pacific HIV Observational Database and is an initiative of TREAT Asia, a program of amfAR and The Foundation for AIDS Research. TAHOD is also supported by the Dutch Ministry of Foreign Affairs through a partnership with Stichting Aids Fonds. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. M Fox was supported by Cooperative Agreement AID 674-A-12-00029 from the United States Agency for International Development (USAID). OK was supported by a PROSPER Ambizione fellowship of the Swiss National Science Foundation.

Acknowledgements: We thank all patients, clinical, management, data entry and support staff in the participating clinics.

ACCEPTED

References

1. World Health O, UNAIDS, UNICEF (2013) Global update on HIV treatment 2013: results, impact and opportunities. Available at http://www.unaids.org/en/media/unaids/contentassets/documents/unaidspublication/2013/20130630_treatment_report_en.pdf (accessed 26 June 2014).
2. World Health O (2012) WHO HIV Drug Resistance Report 2012. Available at http://apps.who.int/iris/bitstream/10665/75183/1/9789241503938_eng.pdf (accessed 26 June 2014).
3. Sigaloff KC, Hamers RL, Wallis CL, Kityo C, Siwale M, et al. (2011) Unnecessary antiretroviral treatment switches and accumulation of HIV resistance mutations; two arguments for viral load monitoring in Africa. *J Acquir Immune Defic Syndr* 58: 23–31. doi:10.1097/QAI.0b013e318227fc34.
4. Keiser O, Orrell C, Egger M, Wood R, Brinkhof MWG, et al. (2008) Public-health and individual approaches to antiretroviral therapy: township South Africa and Switzerland compared. *PLoS Med* 5: e148. doi:10.1371/journal.pmed.0050148.
5. World Health O (2006) Antiretroviral therapy for HIV Infection in adults and adolescents in resource-limited settings: towards universal access. Recommendations for a public health approach. Available at <http://www.who.int/hiv/pub/guidelines/artadultguidelines.pdf> (accessed 26 June 2014).
6. Keiser O, MacPhail P, Boule A, Wood R, Schechter M, et al. (2009) Accuracy of WHO CD4 cell count criteria for virological failure of antiretroviral therapy. *Trop Med Int Health* 14: 1220–1225. doi:10.1111/j.1365-3156.2009.02338.x.
7. Mee P, Fielding KL, Charalambous S, Churchyard GJ, Grant AD (2008) Evaluation of the WHO criteria for antiretroviral treatment failure among adults in South Africa. *AIDS Lond Engl* 22: 1971–1977. doi:10.1097/QAD.0b013e32830e4cd8.
8. Keiser O, Tweya H, Boule A, Braitstein P, Schechter M, et al. (2009) Switching to second-line antiretroviral therapy in resource-limited settings: comparison of programmes with and without viral load monitoring. *AIDS* 23: 1867–1874.
9. World Health O (2013) Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach. Geneva: World Health Organization. Available at http://apps.who.int/iris/bitstream/10665/85321/1/9789241505727_eng.pdf (accessed 26 June 2014).
10. World Health O (2014) March 2014 supplement to the 2013 consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection. Recommendations for a public health approach. Geneva: World Health Organization. Available at http://apps.who.int/iris/bitstream/10665/104264/1/9789241506830_eng.pdf?ua=1 (accessed 26 June 2014).

11. Lynen L, An S, Koole O, Thai S, Ros S, et al. (2009) An algorithm to optimize viral load testing in HIV-positive patients with suspected first-line antiretroviral therapy failure in Cambodia. *J Acquir Immune Defic Syndr* 52: 40–48. doi:10.1097/QAI.0b013e3181af6705.
12. Egger M, Ekouevi DK, Williams C, Lyamuya RE, Mukumbi H, et al. (2012) Cohort Profile: the international epidemiological databases to evaluate AIDS (IeDEA) in sub-Saharan Africa. *Int J Epidemiol* 41: 1256–1264. doi:10.1093/ije/dyr080.
13. Gange SJ, Kitahata MM, Saag MS, Bangsberg DR, Bosch RJ, et al. (2007) Cohort profile: the North American AIDS Cohort Collaboration on Research and Design (NA-ACCORD). *Int J Epidemiol* 36: 294–301. doi:10.1093/ije/dyl286.
14. McGowan CC, Cahn P, Gotuzzo E, Padgett D, Pape JW, et al. (2007) Cohort Profile: Caribbean, Central and South America Network for HIV research (CCASAnet) collaboration within the International Epidemiologic Databases to Evaluate AIDS (IeDEA) programme. *Int J Epidemiol* 36: 969–976.
15. Lawn SD, Harries AD, Anglaret X, Myer L, Wood R (2008) Early mortality among adults accessing antiretroviral treatment programmes in sub-Saharan Africa. *AIDS* 22: 1897–1908.
16. Boulle A, Van Cutsem G, Hilderbrand K, Cragg C, Abrahams M, et al. (2010) Seven-year experience of a primary care antiretroviral treatment programme in Khayelitsha, South Africa. *AIDS* 24: 563–572.
17. Eshun-Wilson I, Plas HV, Prozesky HW, Zeier MD, Nachega J, et al. (2009) Combined antiretroviral treatment initiation during hospitalization: outcomes in South African adults. *J Acquir Immune Defic Syndr* 51: 105–106. doi:10.1097/QAI.0b013e3181963cd4.
18. Ojikutu BO, Zheng H, Walensky RP, Lu Z, Losina E, et al. (2008) Predictors of mortality in patients initiating antiretroviral therapy in Durban, South Africa. *S Afr Med J* 98: 204–208.
19. Fox MP, Maskew M, Macphail AP, Long L, Brennan AT, et al. (2012) Cohort Profile: The Themba Lethu Clinical Cohort, Johannesburg, South Africa. *Int J Epidemiol*. Available at <http://www.ncbi.nlm.nih.gov/pubmed/22434860> (accessed 26 June 2014).
20. Hoffmann CJ, Charalambous S, Thio CL, Martin DJ, Pemba L, et al. (2007) Hepatotoxicity in an African antiretroviral therapy cohort: the effect of tuberculosis and hepatitis B. *AIDS* 21: 1301–1308.
21. Houlihan CF, Bland RM, Mutevedzi PC, Lessells RJ, Ndirangu J, et al. (2011) Cohort profile: Hlabisa HIV treatment and care programme. *Int J Epidemiol* 40: 318–326. doi:10.1093/ije/dyp402.
22. Fatti G, Grimwood A, Bock P (2010) Better antiretroviral therapy outcomes at primary healthcare facilities: an evaluation of three tiers of ART services in four South African provinces. *PLoS One* 5: e12888. doi:10.1371/journal.pone.0012888.
23. Stringer JS, Zulu I, Levy J, Stringer EM, Mwango A, et al. (2006) Rapid scale-up of antiretroviral therapy at primary care sites in Zambia: feasibility and early outcomes. *JAMA J Am Med Assoc* 296: 782–793.
24. Zhou J, Kumarasamy N, Ditangco R, Kamarulzaman A, Lee CK, et al. (2005) The TREAT Asia HIV Observational Database: baseline and retrospective data. *J Acquir Immune Defic Syndr* 38: 174–179.

25. Wood SN (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J R Stat Soc B* 73: 3–36.
26. Wood SN (2003) Thin-plate regression splines. *J R Stat Soc B* 65: 95–114.
27. Liu T, Hogan JW, Wang L, Zhang S, Kantor R (2013) Optimal Allocation of Gold Standard Testing under Constrained Availability: Application to Assessment of HIV Treatment Failure. *J Am Stat Assoc* Epub ahead of print.
28. Lynch S, Ford N, van Cutsem G, Bygrave H, Janssens B, et al. (2012) Public health. Getting HIV treatment to the most people. *Science* 337: 298–300. doi:10.1126/science.1225702.
29. UNAIDS Secretariat (2013) HIV/AIDS diagnostic Technology Landscape. 3rd Edition. Geneva: World Health Organization. Available at http://www.unaids.org/images/marketdynamics/publications/UNAIDS-HIV_Diagnostic_Landscape-3rd_edition.pdf (accessed 26 June 2014).
30. Rowley CF (2014) Developments in CD4 and viral load monitoring in resource-limited settings. *Clin Infect Dis* 58: 407–412. doi:10.1093/cid/cit733.
31. Estill J, Egger M, Blaser N, Vizcaya LS, Garone D, et al. (2013) Cost-effectiveness of point-of-care viral load monitoring of ART in resource-limited settings: Mathematical modelling study. *AIDS*. Available at <http://www.ncbi.nlm.nih.gov/pubmed/23462219> (accessed 26 June 2014).
32. Meintjes G, Boule A, Conradie F, Goemaere E, Hefer E, et al. (n.d.) Guidelines for antiretroviral therapy in adults by the Southern African HIV Clinicians Society. *Afr J HIV Med* 13: 114–133.
33. Phan V, Thai S, Koole O, Menten J, Meheus F, et al. (2013) Validation of a clinical prediction score to target viral load testing in adults with suspected first-line treatment failure in resource-constrained settings. *J Acquir Immune Defic Syndr* 62: 509–516. doi:10.1097/QAI.0b013e318285d28c.
34. Abouyannis M, Menten J, Kiragga A, Lynen L, Robertson G, et al. (2011) Development and validation of systems for rational use of viral load testing in adults receiving first-line ART in sub-Saharan Africa. *AIDS* 25: 1627–1635. doi:10.1097/QAD.0b013e328349a414.
35. Meya D, Spacek LA, Tibenderana H, John L, Namugga I, et al. (2009) Development and evaluation of a clinical algorithm to monitor patients on antiretrovirals in resource-limited settings using adherence, clinical and CD4 cell count criteria. *J Int AIDS Soc* 12: 3. doi:10.1186/1758-2652-12-3.
36. Justice AC, Covinsky KE, Berlin JA (1999) Assessing the generalizability of prognostic information. *Ann Intern Med* 130: 515–524.
37. Mugenyi P, Walker AS, Hakim J, Munderi P, Gibb DM, et al. (2010) Routine versus clinically driven laboratory monitoring of HIV antiretroviral therapy in Africa (DART): a randomised non-inferiority trial. *Lancet* 375: 123–131.
38. Mermin J, Ekwaru JP, Were W, Degerman R, Bunnell R, et al. (2011) Utility of routine viral load, CD4 cell count, and clinical monitoring among adults with HIV receiving antiretroviral therapy in Uganda: randomised trial. *BMJ* 343: d6792. doi:10.1136/bmj.d6792.

39. Bisson GP, Gross R, Bellamy S, Chittams J, Hislop M, et al. (2008) Pharmacy refill adherence compared with CD4 count changes for monitoring HIV-infected adults on antiretroviral therapy. PLoSMed 5: e109.

ACCEPTED

Figure 1. Risk chart for virologic failure based on time on antiretroviral therapy and current CD4 cell count stratified by baseline CD4 (columns) and gender (rows). The area between two lines of the same style contains the patients that are optimally tested given the resources available.

Figure 2. Risk chart for virologic failure based on time on antiretroviral therapy and current CD4 cell count stratified by CD4 count measured 6 months previously (columns) and gender (rows). The area between two lines of the same style contains the patients that are optimally tested given the resources available.

Figure 3. Receiver Operator Curves for Model 1 (upper panels) and Model 2 (lower panels) for the derivation and validation datasets (columns). The curves were computed assuming that no patients, 10%, 20% or 40% of patients were tested for viral load. The numbers indicate the area under the curve (AUC) value of the corresponding curves.

ACCEPTED

Table 1. Baseline characteristics of patients included in the development and validation of risk charts for virologic failure on antiretroviral therapy based on current and baseline CD4 count.

	South Africa (derivation)	South Africa (validation)	Zambia (validation)	Asia-Pacific (validation)
Patients				
Number of patients	31,450 (100%)	16,131 (100%)	7,796 (100%)	1,356 (100%)
Gender				
Female	18,597 (59%)	11,143 (69%)	4,237 (54%)	438 (32%)
Male	12,853 (41%)	4,988 (31%)	3,559 (46%)	918 (68%)
Age (years)				
Median (IQR)	36 (30 – 42)	35 (30 – 42)	36 (31 – 43)	35 (31 – 43)
16 - 29	6,728 (21%)	3,751 (23%)	1,370 (18%)	279 (21%)
30 - 39	13,899 (44%)	7,241 (45%)	3,599 (46%)	615 (45%)
40 - 49	7,705 (24%)	3,725 (23%)	2,021 (26%)	324 (24%)
>= 50	3,118 (10%)	1,414 (9%)	806 (10%)	138 (10%)
CD4 count at start of ART (cells/μl)				
Median (IQR)	111 (50 – 174)	129 (72 – 174)	114 (58 – 180)	95 (31 – 184)
< 50	7,800 (25%)	2,611 (16%)	1,621 (21%)	463 (34%)
50 - 99	6,465 (21%)	3,321 (21%)	1,788 (23%)	226 (17%)
100 - 199	12,581 (40%)	8,400 (52%)	2,973 (38%)	375 (28%)
200 - 349	4,593 (15%)	1,792 (11%)	1,409 (18%)	292 (22%)
Year of starting ART				
Median (IQR)	2006 (2005 – 2007)	2008 (2007 – 2009)	2006 (2005 – 2007)	2004 (2002 – 2005)
Follow up time (years)				
Median (IQR)	1.92 (1.09 – 3.04)	1.72 (1.06 – 2.70)	2.50 (1.64 – 3.32)	3.39 (2.03 – 4.58)
Total	68,611	32,005	19,951	4,375
Laboratory values				
No. of triplets analysed*	125,590 (100%)	46,997 (100%)	16,652 (100%)	8,498 (100%)
No. with virologic failure	20,879 (17%)	10,642 (23%)	2,904 (17%)	828 (11%)
No. of imputed CD4 counts	12,045 (10%)	14,355 (31%)	5,980 (36%)	532 (6%)
No. of imputed VL measurements	11,972 (10%)	14,468 (31%)	4,815 (29%)	3,214 (38%)

*: Baseline CD4 count and CD4 count and VL measured at the same time during follow-up.

VL, viral load

Table 2. Accuracy of prediction of virologic failure in derivation and validation cohorts. Results are shown for no viral load testing and for the testing of 20% or 40% of patients, using optimal rules for the range of patients tested based on the predicted probability of virologic failure.

	Current and baseline CD4 count				Current CD4 count and CD4 count measured 3 months previously			
	(Model 1)				(Model 2)			
	South Africa (derivation)	South Africa (validation)	Zambia (validation)	Asia-Pacific (validation)	South Africa (derivation)	South Africa (validation)	Zambia (validation)	Asia-Pacific (validation)
0% viral load testing								
	Probability cut off: 0.38				Probability cut off: 0.36			
PPV	61%	48%	35%	37%	56%	49%	28%	29%

NPV	88%	81%	87%	91%	88%	79%	89%	92%
Sensitivity	33%	24%	43%	25%	24%	19%	29%	18%
Specificity	96%	92%	83%	95%	97%	94%	89%	95%
10% viral load testing								
	Probability range tested: 0.29 - 0.55				Probability range tested: 0.26 - 0.56			
PPV	87%	79%	64%	73%	89%	82%	71%	78%
NPV	90%	84%	91%	93%	90%	82%	92%	93%
Sensitivity	44%	35%	55%	37%	36%	30%	43%	29%
Specificity	99%	97%	93%	98%	99%	98%	97%	99%
% tested	10%	14%	22%	12%	10%	13%	22%	12%
20% viral load testing								
	Probability range tested: 0.22 - 0.64				Probability range tested: 0.2 - 0.67			
PPV	94%	91%	80%	88%	96%	93%	88%	93%
NPV	92%	87%	93%	94%	92%	84%	94%	94%
Sensitivity	56%	49%	68%	49%	50%	42%	56%	41%
Specificity	99%	99%	96%	99%	100%	99%	99%	100%
% tested	20%	26%	40%	24%	20%	24%	39%	24%
40% viral load testing								
	Probability range tested: 0.14 - 0.76				Probability range tested: 0.13 - 0.94			
PPV	98%	98%	93%	96%	100%	100%	100%	100%
NPV	95%	92%	96%	97%	95%	89%	96%	96%
Sensitivity	74%	68%	82%	71%	69%	62%	73%	63%
Specificity	100%	100%	99%	100%	100%	100%	100%	100%
% tested	40%	49%	65%	45%	40%	45%	62%	44%

Figure 1. Risk chart for virologic failure based on time on antiretroviral therapy and current CD4 cell count stratified by baseline CD4 (columns) and gender (rows). The area between two lines of the same style contains the patients that are optimally tested given the resources available.

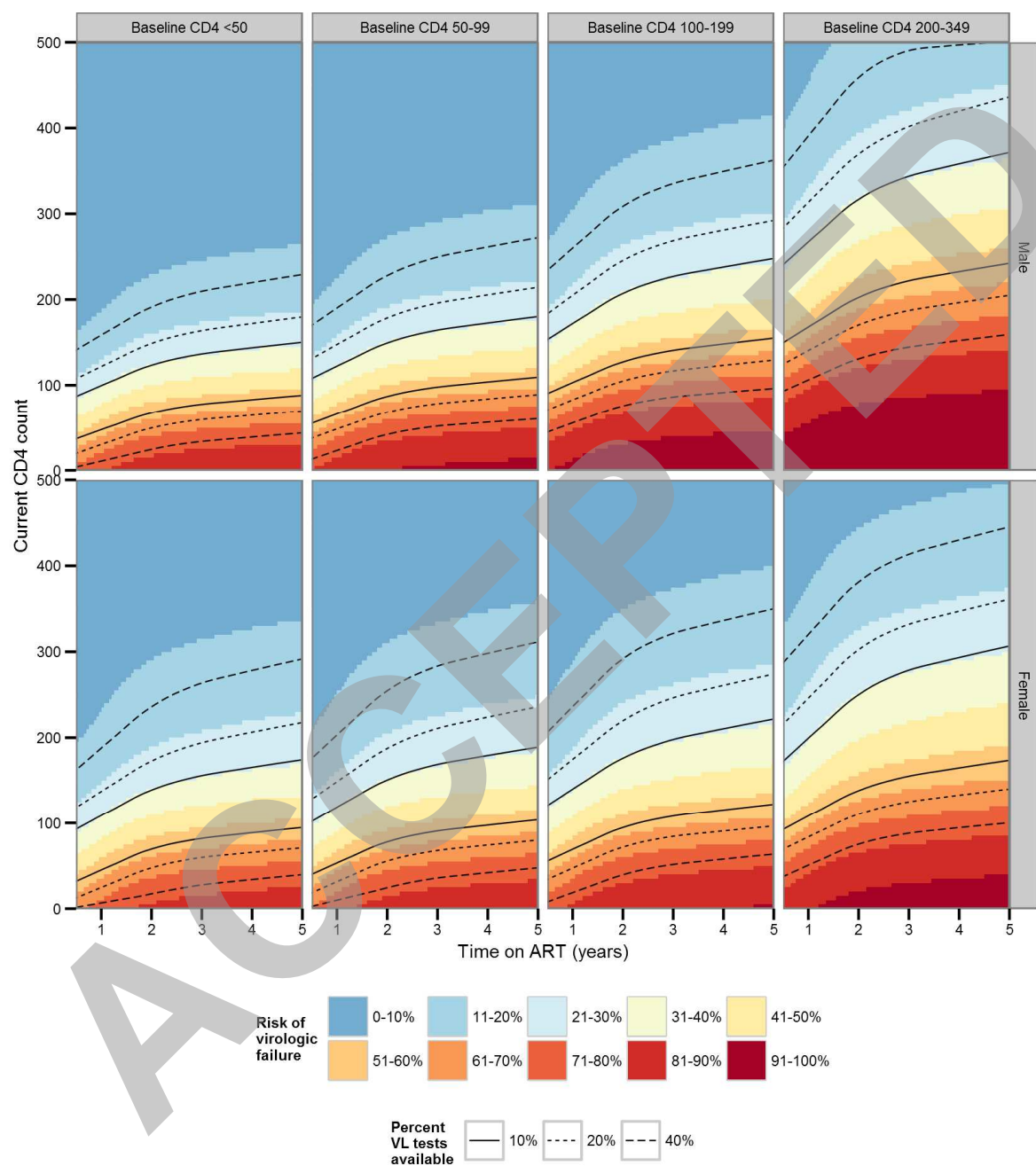


Figure 2. Risk chart for virologic failure based on time on antiretroviral therapy and current CD4 cell count stratified by CD4 count measured 6 months previously (columns) and gender (rows). The area between two lines of the same style contains the patients that are optimally tested given the resources available.

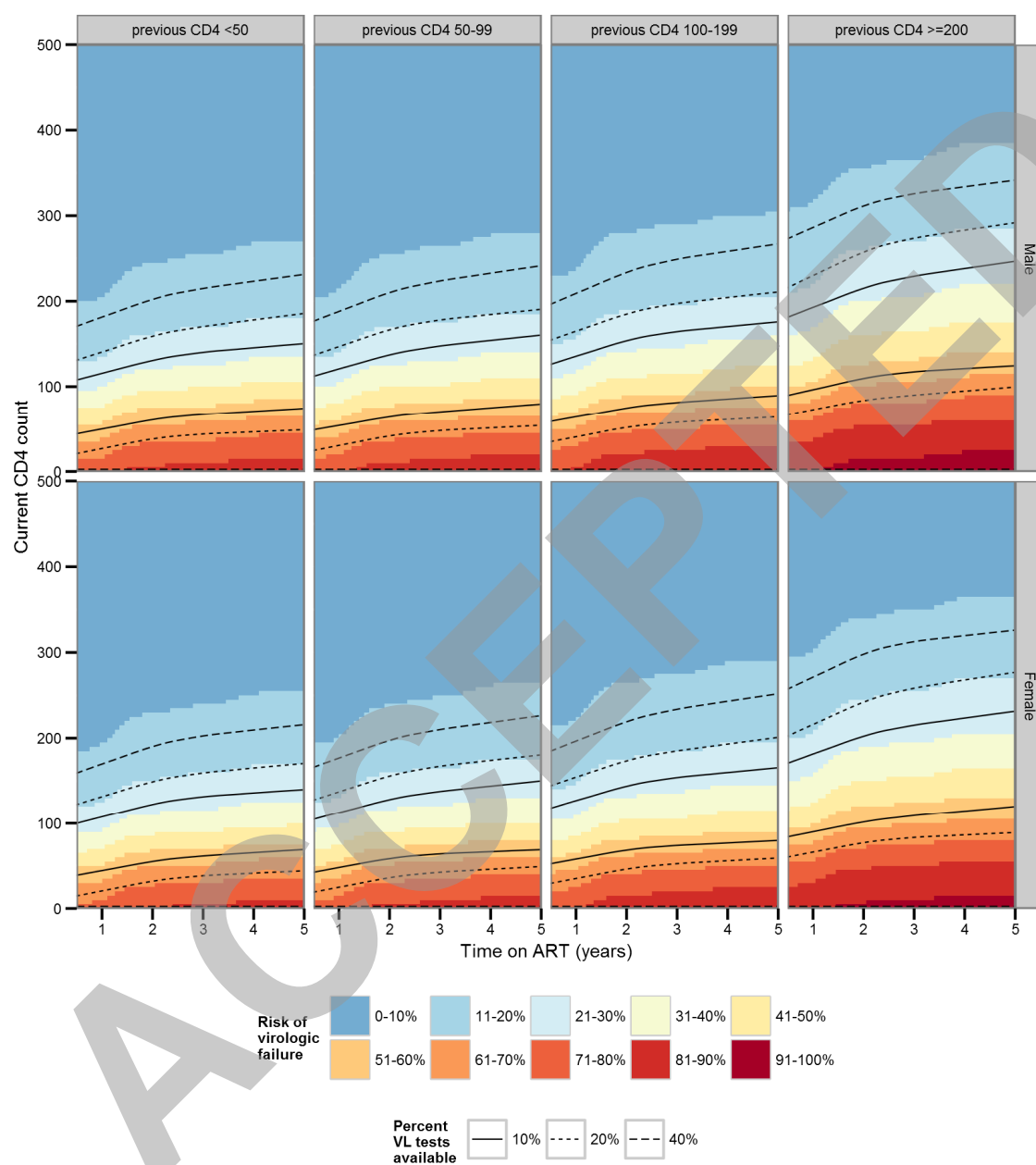


Figure 3. Receiver Operator Curves for Model 1 (upper panels) and Model 2 (lower panels) for the derivation and validation datasets (columns). The curves were computed assuming that no patients, 10%, 20% or 40% of patients were tested for viral load. The numbers indicate the area under the curve (AUC) value of the corresponding curves.

